

⑫ 公開特許公報(A)

平3-40177

⑤Int.Cl.⁵

G 06 F 15/62
15/70
G 10 L 3/00
5/06

識別記号

3 8 0
4 6 0 B
3 0 1 Z
B

庁内整理番号

8419-5B
9071-5B
8842-5D
8842-5D

④公開 平成3年(1991)2月20日

審査請求 未請求 請求項の数 3 (全3頁)

⑥発明の名称 音声認識装置

②特 願 平1-176276

②出 願 平1(1989)7月7日

⑦発 明 者 伴 田 俊 二 長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内

⑦出 願 人 セイコーエプソン株式会社 東京都新宿区西新宿2丁目4番1号

⑦代 理 人 弁理士 鈴木 喜三郎 外1名

明 細 書

1. 発明の名称

音声認識装置

2. 特許請求の範囲

(1) 音声処理に加え、発生音と唇の動きとの相関による情報を画像処理手法を用いて付加することにより認識を行うことを特徴とする音声認識装置。

(2) 唇形状データ及び音声特徴の記号化データの両方を用いて認識処理部で音韻認識処理することを特徴とする請求項1記載の音声認識装置。

(3) 唇形状の知識データベース、発声音と唇の動きとの相関データベース及び言語処理データベースとを含むことを特徴とする請求項1、または請求項2記載の音声認識装置。

3. 発明の詳細な説明

(産業上の利用分野)

本発明は音声入力、音声通信等に用いられる音声認識装置に関する。

(発明の概要)

本発明は入力音声の処理のみならず、発生音と唇の動きとの相関による情報を画像処理手法を用いて付加することにより認識を行う音声認識装置である。

(従来技術)

一般に音声理解システムは、音声の物理的な特徴である音響情報と言語情報に代表される高次情報とを用いて、発話の構文あるいは意味的な制約情報を付加して音声の認識を行っている。

システムの流れとしては、入力音声を受けてからの音響処理部、文節候補生成部及び言語処理部等から構成されるが、音響処理部においては音韻の位置と種類とを知識工学の手法を用いてシュミレートし、音韻単位の認識を行う。もう少し詳しく述べると、入力音声は音響処理部の特徴抽出部で分析及び特徴抽出され、記号化されたデータは知識データベースを持つ音響処理部の認識処理部

へ送られ、ルール表現された知識を適用することにより音韻境界と音韻の認識が行われる。

音韻処理により得られた文節単位の音韻列は、文節候補生成部において文節内の構文規則を考慮した上で単語辞書とのマッチングが行われ、該当する文節候補が生成される。

言語処理部では、文節候補生成部において生成された文節候補列を入力として、構文、意味及び文脈等の知識を用いて正しい文節列を同定する。
〔発明が解決しようとする課題〕

しかし、従来の認識手法においては音声のみによる特徴抽出、認識処理のため、不特定話者の音声に対しては認識が困難であり、従って膨大な知識としてのデータベースを抱える必要があった。

さらには、膨大なデータベースを用いたとしても、不特定話者の音韻特徴は非常に異なっており、単純な母音と子音とを組み合わせた程度の短音認識が具現化されている程度である。さらには膨大なデータベースを検索するための時間を要し、リアルタイムで処理することも困難であった。

-3-

システムの一つの例と同じであり、従来の技術の所で概要を述べてある。本発明ではさらなる付加処理として、顔画像をカメラ等で取り込み（入力顔画像（1））、唇画像分離（5）によりカメラから取り込まれた画像のうち唇部分だけを分離・抽出し、画像処理部（6）で、例えばワイヤフレーム法（領域を3角形等の要素に分割し、それぞれの交点の情報を、数点の基準点からの変化の情報として得、画像の動きを検出する。）等の方法で唇の形状変化を解析した後、ワイヤフレーム法に基づく唇形状の情報をデータとして持っているデータベース（4）の内容を検索することにより、得られた唇形状データ列を認識処理部（8）に送る。認識処理部（8）においては、音声データベースとともに発声音と唇の動きとの関連データを持っているデータベース（10）の内容を検索することにより、特徴抽出部（7）から得られた音声特徴の記号化データとともに音韻の認識を行い、確定音韻列として文節候補生成部（9）に送る。文節候補生成部（9）では、文節内の構文規

そこで本発明では、認識手法の一部として音声以外の情報を付加することによって、認識率の向上と高速化を図ることを目的とする。

〔課題を解決するための手段〕

本発明の音声認識装置は、

（1）音声処理に加え、発声音と唇の動きとの相関による情報を画像処理手法を用いて付加することにより認識を行うことを特徴とする。

（2）唇形状データ及び音声特徴の記号化データの両方を用いて認識処理部で音韻認識処理する。

（3）唇形状の知識データベース、発声音と唇の動きとの相関データベース及び言語処理データベースとを含むことを特徴とする。

〔実施例〕

第1図は本発明の実施例の処理フローを示すブロック図である。

入力音声（2）、特徴抽出部（7）、認識処理部（8）、文節候補生成部（9）、データベース（10）、単語辞書（11）、言語処理部（12）及びデータベース（13）等の流れは、従来の音声理解

-4-

則を考慮した上で単語辞書とのマッチングが行われ、該当する文節候補が生成される。タイミング発生回路（3）は、入力音声と唇の動きとの同期データを、唇形状データ列と音声特徴の記号化データとの取り込み部である認識処理部に送るためのものであり、本図面では省略してあるが、唇の動きと発声音とのタイミングについては全て中央処理装置のコントロール下に置かれる。文節候補生成部（9）からの文節候補列は言語処理部（12）に送られ、言語処理部（12）は言語処理データベース（13）の内容を検索することにより、構文、意味及び文脈等の知識を用いて正しい文節列を同定し出力する。

〔発明の効果〕

以上本発明によれば、唇の動きは話者には余り依存しないため、一つあるいは数種のANDをとった典型的な発声音と唇の動きとのデータベースを持てばよく、膨大な知識としてのデータベースを抱える必要がない。さらには同じ様な理由から、認識対象者が非常に独特な音声特徴を持つ不特定

話者であっても、唇の動きによる認識フローの付加により音韻境界と音韻とを認識することは容易であり、データベース量の少ないより身軽なシステムにおいても、短音のみならず連続発声音の認識が可能になる。さらには、データベース量の少ない分だけ検索に要する時間が少なくなり、よりリアルタイムな音声理解システムが得られる。

従って、本発明はコンピューターの入力装置、自動翻訳装置等の音声通信機器、音声入力のワードプロセッサ、等のよりコンパクト性が必要とされるシステムに応用できる。

4. 図面の簡単な説明

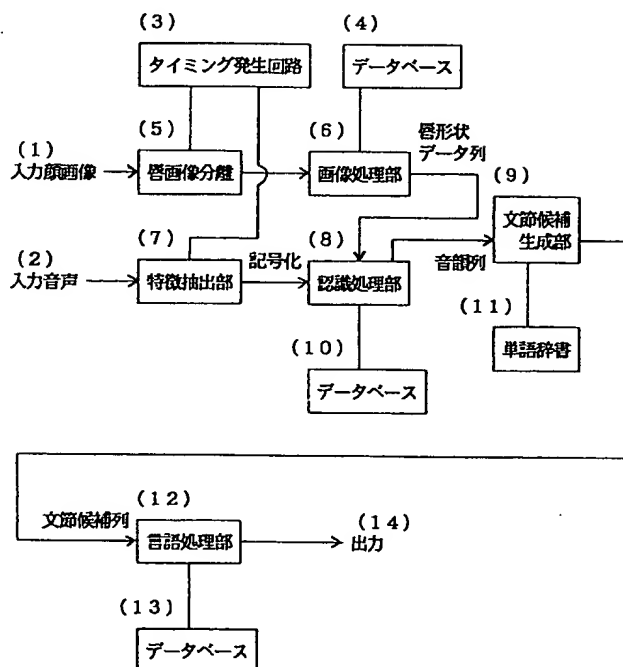
第1図は本発明の実施例の処理フローを示すブロック図。

以 上

出願人 セイコーエプソン株式会社

代理人 弁理士 鈴木 喜三郎 他1名

-7-



第 1 図